J|A|C|S ARTICLES

Published on Web 07/24/2009

Highly Conserved Histidine Plays a Dual Catalytic Role in Protein Splicing: A pK_a Shift Mechanism

Zhenming Du,[†] Philip T. Shemella,[‡] Yangzhong Liu,[§] Scott A. McCallum,[†] Brian Pereira,[||,⊥] Saroj K. Nayak,[‡] Georges Belfort,[⊥] Marlene Belfort,[||] and Chunyu Wang*,[†]

Biology Department, Center for Biotechnology and Interdisciplinary Studies, Rensselaer Polytechnic Institute, Troy, New York 12180, Department of Physics, Applied Physics, and Astronomy, Rensselaer Polytechnic Institute, Troy, New York 12180, Department of Chemistry, University of Science and Technology of China, Hefei, Anhui, P.R. China 230026, Wadsworth Center, New York State Department of Health, Center for Medical Sciences, 150 New Scotland Avenue, Albany, New York 12208, and Department of Chemical and Biological Engineering, Center for Biotechnology and Interdisciplinary Studies, Rensselaer Polytechnic Institute, Troy, New York 12180

Received May 27, 2009; E-mail: wangc5@rpi.edu

Abstract: Protein splicing is a precise autocatalytic process in which an intein excises itself from a precursor with the concomitant ligation of the flanking sequences. Protein splicing occurs through acid−base catalysis in which the ionization states of active site residues are crucial to the reaction mechanism. In inteins, several conserved histidines have been shown to play important roles in protein splicing, including the most conserved "B-block" histidine. In this study, we have combined NMR pK_a determination with quantum mechanics/molecular mechanics (QM/MM) modeling to study engineered inteins from *Mycobacterium tuberculosis* (*Mtu*) RecA intein. We demonstrate a dramatic pK_a shift for the invariant B-block histidine, the most conserved residue among inteins. The B-block histidine has a pK_a of 7.3 ± 0.6 in a precursor and a pK_a of <3.5 in a spliced intein. The pK_a values and QM/MM data suggest that the B-block histidine has a dual role in the acid−base catalysis of protein splicing. This histidine likely acts as a general base to initiate splicing with an acyl shift and then as a general acid to cause the breakdown of the scissile bond at the N-terminal splicing junction. The proposed pK_a shift mechanism accounts for the biochemical data supporting the essential role for the B-block histidine and for the near absolute sequence conservation of this residue.

Introduction

Protein splicing is an autocatalytic process in which an in-frame protein fusion, called an intein, is excised from the precursor protein with the concomitant ligation of the two flanking polypeptides, the N- and C-exteins (Figure 1).[1] Described by Tom Muir as "Nature's gift to the protein chemist",[2] inteins are widely used in protein engineering, protein labeling, protein purification, and control of protein functions.[3−6]

The four steps of the protein splicing pathway have been well documented (Figure 1).[1,7] In the first step, *N−X acyl shift (X = S or O)*, the side chain nucleophile of the first residue of intein (side chain S atom of a cysteine or O atom of a serine) attacks the carbonyl of the last residue of the N-extein,[8,9] resulting in a linear ester intermediate. In the second step, *transesterification*, the nucleophile at the downstream splice junction attacks the linear ester. The N-extein is transferred to the side chain of the attacking nucleophile, forming a branched ester intermediate.[10−13] In the third step, *asparagine cyclization and C-terminal cleavage*, the last residue of the intein, an asparagine, cyclizes. This is coupled to the cleavage of the branched ester and the release of the excised intein with an aminosuccinimide residue and the exteins joined by an ester bond.[9,10,14,15] In the fourth step, *X−N*

[†] Biology Department, Center for Biotechnology and Interdisciplinary Studies, Rensselaer Polytechnic Institute.
[‡] Department of Physics, Applied Physics, and Astronomy, Rensselaer Polytechnic Institute.
[§] University of Science and Technology of China.
[||] New York State Department of Health.
[⊥] Department of Chemical and Biological Engineering, Center for Biotechnology and Interdisciplinary Studies, Rensselaer Polytechnic Institute.

(1) Paulus, H. *Chem. Soc. Rev.* **1998**, *27*, 375–386.
(2) Belfort, M. In *Homing Endonucleases and Inteins*; Belfort, M., Derbyshire, V., Stoddard, B. L., Wood, D. W., Eds.; Springer: Berlin Heidelberg, 2005; Vol. 16, pp 1−9.
(3) Blaschke, U. K.; Silberstein, J.; Muir, T. W. *Methods Enzymol.* **2000**, *328*, 478–96.
(4) Mootz, H. D.; Blum, E. S.; Tyszkiewicz, A. B.; Muir, T. W. *J. Am. Chem. Soc.* **2003**, *125*, 10561–9.
(5) Tyszkiewicz, A. B.; Muir, T. W. *Nat. Methods* **2008**, *5*, 303–5.
(6) Wood, D. W.; Derbyshire, V.; Wu, W.; Chartrain, M.; Belfort, M.; Belfort, G. *Biotechnol. Prog.* **2000**, *16*, 1055–63.
(7) Saleh, L.; Perler, F. B. *Chem. Record* **2006**, *6*, 183–193.
(8) Shao, Y.; Xu, M. Q.; Paulus, H. *Biochemistry* **1996**, *35*, 3810–5.
(9) Chong, S.; Shao, Y.; Paulus, H.; Benner, J.; Perler, F. B.; Xu, M. Q. *J. Biol. Chem.* **1996**, *271*, 22159–68.
(10) Xu, M. Q.; Comb, D. G.; Paulus, H.; Noren, C. J.; Shao, Y.; Perler, F. B. *Embo J.* **1994**, *13*, 5517–22.
(11) Xu, M. Q.; Southworth, M. W.; Mersha, F. B.; Hornstra, L. J.; Perler, F. B. *Cell* **1993**, *75*, 1371–7.
(12) Chong, S.; Xu, M. Q. *J. Biol. Chem.* **1997**, *272*, 15587–90.
(13) Nichols, N. M.; Benner, J. S.; Martin, D. D.; Evans, T. C., Jr. *Biochemistry* **2003**, *42*, 5301–11.
(14) Shao, Y.; Xu, M. Q.; Paulus, H. *Biochemistry* **1995**, *34*, 10844–50.

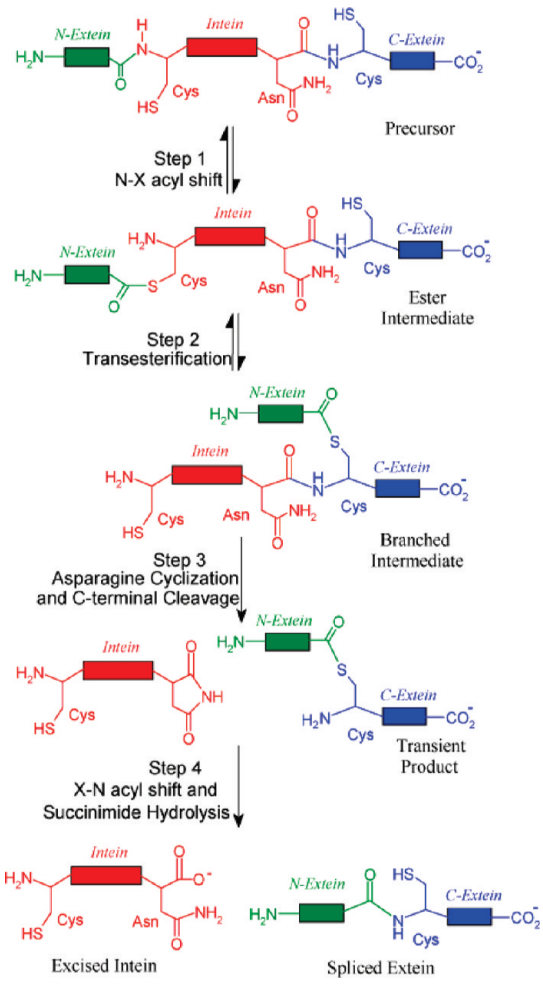

Figure 1. Four steps of protein splicing. During the four-step reaction, the intein (shown in red) is excised from the N-extein (shown in green) and C-extein (shown in blue), while the N-extein and C-exteins are ligated to form the mature protein.

acyl shift and succinimide hydrolysis, the aminosuccinimide hydrolyzes and the ester linking the two exteins rearranges to form a peptide bond.[14,16] For most of these steps, the mechanistic details are still lacking at the atomic level. For example, it is not known which residue activates the side chain nucleophile to initiate the N−X shift in the first step of splicing.

The intein sequence contains conserved sequence blocks (Figure 2). Block A residues almost always start with a cysteine or a serine (Figure 2), providing a nucleophile for the N−X acyl shift (Figure 1). Block B contains the TXXH motif, where both T and H play important roles for the first step of protein splicing.[17−19] The F block often contains an aspartate[20] and a histidine that modulates C-terminal cleavage.[21,22] The G block has a penultimate histidine and a C-terminal asparagine critical for C-terminal cleavage.[9,10,14,15] The first residue of the C-extein is a C, S, or T (Figure 2), serving as the nucleophile for transesterification, the second step of the protein splicing reaction (Figure 1).

The B-block histidine is the most conserved residue in all intein sequences (Figure 2).[23,24] Mutagenesis studies have shown that the B-block histidine plays an essential role in splicing by catalyzing the N−X acyl shift.[18,25,26] Crystal structures of inteins have shown that that the δ1 nitrogen of the B-block histidine is close to the amide nitrogen of the first residue of the intein, suggesting that the B-block histidine may promote N−X acyl shift by protonating the leaving group.[7,17,19,27−30] The early crystal structure of the GyrA intein showed the scissile bond at the upstream splice junction to be in an unusual *cis* conformation, which may facilitate the N−X acyl shift by ground state destabilization.[17] Consistent with this hypothesis, a clever NMR study revealed an unusually low $^{1}J_{NC'}$ (12 Hz) of the N-terminal scissile bond,[31] which reverted to a normal J-coupling with an alanine mutation of the B-block histidine. These results indicate that the B-block histidine may destabilize the ground state in a precursor. However, later crystal structures demonstrated a variety of conformations for the scissile bond, ranging from *trans*[19,22,29,32] to distorted *trans* conformation.[28] Thus, the precise catalytic role of the invariant B-block histidine remains to be defined.

The ionization states of active site residues are crucial to the mechanism of acid−base catalysis in enzymes. Many catalytic residues have an elevated or depressed pK_a.[33] Solution NMR is the ideal tool for site-specific pK_a measurement in enzymes. Although several NMR studies of inteins have been published,[31,34,35] the pK_a values of conserved histidines have not been determined in any intein. In this study, NMR pK_a measurement and quantum mechanics/molecular mechanics (QM/MM) modeling have been carried out for engineered and minimized (139 aa residues) *Mtu* RecA inteins, ΔΔIhh-SM (splicing mutant with a V67L mutation) and ΔΔIhh-CM (cleavage mutant with V67L/D422G mutations).[20] These inteins have been minimized by the deletion of the dispensable endonuclease domain[20] and the replacement of a long, disordered loop with a short loop from the homologous hedgehog (hh) protein.[36] The V67L mutation globally promotes the splicing reaction, while the additional D422G mutation enhances C-terminal cleavage.[20] X-ray crystallography studies of these inteins demonstrate a typical HINT (hedgehog intein) horseshoe fold and close proximity of active site residues (Figure 3).[30]

(15) Xu, M. Q.; Perler, F. B. *Embo J.* **1996**, *15*, 5146–53.
(16) Shao, Y.; Paulus, H. *J. Pept. Res.* **1997**, *50*, 193–8.
(17) Klabunde, T.; Sharma, S.; Telenti, A.; Jacobs, W. R., Jr.; Sacchettini, J. C. *Nat. Struct. Biol.* **1998**, *5*, 31–6.
(18) Mizutani, R.; Anraku, Y.; Satow, Y. *J. Synchrotron Radiat.* **2004**, *11*, 109–112.
(19) Mizutani, R.; Nogami, S.; Kawasaki, M.; Ohya, Y.; Anraku, Y.; Satow, Y. *J. Mol. Biol.* **2002**, *316*, 919–29.
(20) Wood, D. W.; Wu, W.; Belfort, G.; Derbyshire, V.; Belfort, M. *Nat. Biotechnol.* **1999**, *17*, 889–92.
(21) Ghosh, I.; Sun, L.; Xu, M. Q. *J. Biol. Chem.* **2001**, *276*, 24051–8.
(22) Ding, Y.; Xu, M. Q.; Ghosh, I.; Chen, X.; Ferrandon, S.; Lesage, G.; Rao, Z. *J. Biol. Chem.* **2003**, *278*, 39133–42.
(23) Perler, F. B. *Nucleic Acids Res.* **2002**, *30*, 383–384.
(24) Pietrokovski, S. *Protein Sci.* **1994**, *3*, 2340–50.
(25) Kawasaki, M.; Nogami, S.; Satow, Y.; Ohya, Y.; Anraku, Y. *J. Biol. Chem.* **1997**, *272*, 15668–15674.
(26) Ghosh, I.; Sun, L.; Xu, M.-Q. *J. Biol. Chem.* **2001**, *276*, 24051–24058.
(27) Duan, X.; Gimble, F. S.; Quiocho, F. A. *Cell* **1997**, *89*, 555–64.
(28) Poland, B. W.; Xu, M. Q.; Quiocho, F. A. *J. Biol. Chem.* **2000**, *275*, 16408–13.
(29) Sun, P.; Ye, S.; Ferrandon, S.; Evans, T. C.; Xu, M. Q.; Rao, Z. *J. Mol. Biol.* **2005**, *353*, 1093–105.
(30) Van Roey, P.; Pereira, B.; Li, Z.; Hiraga, K.; Belfort, M.; Derbyshire, V. *J. Mol. Biol.* **2007**, *367*, 162–173.
(31) Romanelli, A.; Shekhtman, A.; Cowburn, D.; Muir, T. W. *Proc. Natl. Acad. Sci. U.S.A.* **2004**, *101*, 6397–402.
(32) Anraku, Y.; Mizutani, R.; Satow, Y. *IUBMB Life* **2005**, *57*, 563–74.
(33) Fersht, A. *Structure and Mechanism in Protein Science: A guide to enzyme catalysis and protein folding*; WH Freeman and Company: New York, 1999.
(34) Johnson, M. A.; Southworth, M. W.; Herrmann, T.; Brace, L.; Perler, F. B.; Wuthrich, K. *Protein Sci.* **2007**, *16*, 1316–28.

```
                1 A-block          20         30         40
Mtu RecA        CLAEGTRIFDPVTGTTHRIEDVVDGRKPIHVVAAAKDGTLHARPVVSWFD
Ssp DnaB        CISGDSLISLASTGKRVSIKDLL-DEKDFEIWAINEQMKLESAKVSRVFM
Hedgehog        CFTPESTALLE-SGVRKPLGEL---SIGDRVLSXTANGQAVYSEVILFDR
Ssp DnaE        CLSFGTEILTV-EYGPLPIGKIVSEEINCSVYSVDPEGRVYTQAIAQWHD
Tko Pol-2       SILPEEWLPVLEEVHFVRIGELIDRM-GLEVPSFNRRNKAELKRVKALIR
Mxe GyrA        CITGDALVALP-EGESVRIADIVPGAIDLKVLDR----HGNPVLADRLFH
Mja KlbA        ALAYDEPIYLS-DGNIINIGEFVDKF-NIYIKSFNKSLIIEDKRILRVWR
Sce VMA         CFAKGTNVLMA-DGSIECIENI---EVGNKVMGK----DGRPREVIKLP-

                51        60       B-block     80         90
Mtu RecA        QGTRDVIGLRIAGGAILWATPDHKVLTEYGWRAAGELRKGDRVAVRDVET
Ssp DnaB        TGKKLVYILKTRLGRTIKATANHRFLTIDGWKRLDELSLKEHIALPRK--
Hedgehog        NLXQNFVQLHTDGGAVLTVTPAHLVSVWQTFVFADRIEEKNQVLVRDVET
Ssp DnaE        RGEQEVLEYELEDGSVIRATSDHRFLTTDQLLAIEEIF-QLDLLTLEIKQ
Tko Pol-2       HDYSKVYTIRLKSGRRIKITSGHSLFSVRVEVTGDELKPGDLVAVPRLLL
Mxe GyrA        SGEHPVYAVRTVEGLRVTGTANHPLLCLVLWKLIDEIKPGDYAVIQRS--
Mja KlbA        KKYGKLIKITTKNRREITLTHDHPVYISVLEINAEMVKVGDYIYIPK--N
Sce VMA         RGRETMYSVVQKELLKFTCNATHELVVREWTIEARDLRKATYQTYAPRSL

                403     410        420        430        440
                                   F-block          G-block
Mtu RecA        GELRYSVIREVLPTRRARTFDLEVEELHTLVAEGVVVHNC
Ssp DnaB        -DISWDSIVSITETGVEEVFDLTVPGPHNFVANDIIVHNS
Hedgehog        GELRPQRVVKVGSVRSKGVVAPLTR-EGTIVVNSVAASC-
Ssp DnaE        TTIKXVKVIGRRSLGVQRIFDIGLPQDHNFLLNGAIAANC
Tko Pol-2       NDVVLDRVESVDVEDYGYVYDLSVEDNENFLVGLVYAHNS
Mxe GyrA        -RFYYAKVASVTDAGVQPVYSLRVDTDHAFITNGFVSHNT
Mja KlbA        NTINLDEVIKVETVDYGHIYDLTVEDNHTYIAGGFAVSNC
Sce VMA         GECRGFYF-ELQELKEDDYYGITLSSDHQFLLNQVVVHNC
```

Figure 2. B-block histidine is the most conserved residue in the intein sequences. Structure-based multiple sequence alignment of ΔΔIhh-SM and other inteins was achieved the using DALI server. ΔΔIhh-SM is an engineered and minimized *Mtu* RecA intein (see text). The locations of the conserved blocks A, B, F, and G are indicated above the sequences and the corresponding residues are colored in yellow. The conserved histidines are colored in magenta while other key residues are colored in cyan. H73, H429, and H439 are the conserved B-block, F-block, and penultimate histidines in ΔΔI-hh-SM, respectively. The first C-extein residue is colored in green. The gap (▼) in the residue numbering for ΔΔIhh-SM results from the deletion of the endonuclease domain of *Mtu* RecA intein.

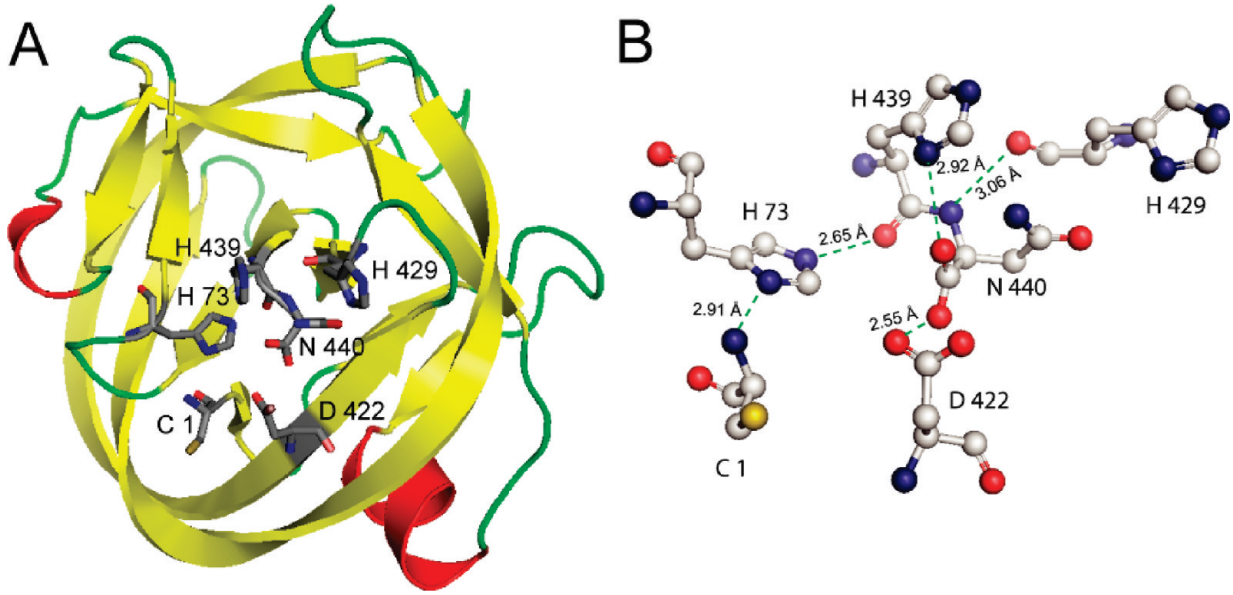

Figure 3. Crystal structure of a minimized *Mtu* RecA intein, ΔΔIhh-SM (pdb code 2IN0). (A) Overall horseshoe fold of HINT domains. Conserved residues, C1, H73, D422, H429, H439, and N440, are shown in stick representation. (B) Hydrogen bond network between conserved residues.

In this study, we show that in intein-mediated protein splicing, the most conserved residue, the B-block histidine, experiences a large pK_a shift and has a strikingly low pK_a in the spliced intein. The pK_a values of the B-block histidine and QM/MM modeling indicate that this histidine likely acts first as a general base with a high pK_a then as a general acid with a low pK_a in the first step of protein splicing.

Materials and Methods

In Vivo* Splicing and Cleavage Assay.** ***Escherichia coli. JM109 cells transformed with pMIC which encodes a fusion protein of maltose binding protein (MBP, M), ΔΔIhh-SM (I) and a small C-extein (C-terminal domain of I-TevI) were grown in 6 mL cultures to an OD_{600} of 0.5 −0.6 at 37 °C in Luria−Bertani (LB) medium. Isopropyl β-D-thiogalactoside (IPTG) was added to a final concentration of 1 mM and the growth was continued for 3 h during

(35) Oeemig, J. S.; Aranko, A. S.; Djupsjobacka, J.; Heinamaki, K.; Iwai, H. *FEBS Lett.* **2009**, *583*, 1451–6.
(36) Hiraga, K.; Derbyshire, V.; Dansereau, J. T.; Van Roey, P.; Belfort, M. *J. Mol. Biol.* **2005**, *354*, 916–926.

which the expressed fusion proteins underwent splicing, C-terminal and/or N-terminal cleavage (Figure S1, Supporting Information). Cells were collected by centrifugation and lysed. The supernatant from the lysate was run on SDS-PAGE to assess the efficiency of protein splicing, C- and/or N-terminal cleavage based on the amount of the spliced and cleaved products. These products have been confirmed by Western blot with respective antibodies.[30]

Protein Overexpression, Purification, and NMR Sample Preparation. NMR samples of ΔΔIhh-SM, ΔΔIhh-CM, and ΔΔIhh-CM with a V67L mutation were prepared as described previously.[30,37] The NI intein precursor is comprised of the intein itself (ΔΔIhh-SM, 139 residues, D24G, C1A), flanked by 17 N-extein residues (MAHHHHHHVGTGSNADP). The C1A mutation was introduced to prevent N-terminal cleavage. The NI precursor was expressed in the *E. coli* strain ER2566. Cells were grown at 37 °C and induced with 1 mM IPTG at an OD_{600} of 0.3−0.4 at 20 °C, then grown for another 20−24 h at 20 °C. Cell pellets were lysed by sonication in 50 mM Tris buffer, pH 8.0, with 0.5 M NaCl and 20 mM imidazole, and purified by affinity chromatography on Ni-NTA agarose (GE Healthcare) followed by gel filtration chromatography on a HiLoad 16/60 Superdex 200 preparation grade column (GE Healthcare). Isotope labeling was accomplished by growing cultures in M9 minimal medium containing either 1 g/L $^{15}NH_4Cl$ for a uniformly ^{15}N-labeled sample, or 1 g/L $^{15}NH_4Cl$ and 1 g/L $^{13}C_6$-D-glucose (Cambridge Isotope Laboratories) for a uniformly ^{15}N,^{13}C-labeled sample. Both procedures yielded ∼20 mg of pure protein from 1 L of culture. The pure protein was concentrated and exchanged by ultrafiltration with Amicon Ultra Centrigugal Filter Devices (Millipore Cooperation) (10-kDa molecular weight cutoff) into 50 mM sodium phosphate buffer, pH 7.0, with 100 mM NaCl and 1 mM NaN_3. The final protein concentration for NI was 1 mM with a volume of 500 μL.

NMR Resonance Assignment. All NMR experiments were carried out at 25 °C on either a Bruker 800 MHz (^1H) or 600 MHz (^1H) spectrometer, each equipped with a triple-resonance cryogenic probe. The complete assignment of ΔΔIhh-CM has been published[37] and the assignment of ΔΔIhh-SM have been carried out in a similar manner. The assignments of histidines are based on the (Hβ)Cβ(CγCδ)Hδ spectrum (see Supporting Information Figures S2−S4), which connects backbone Cβ to Hδ in the imidazole ring.[38]

pH Titration. For pH titration, we started with a pH 7 sample, which was split into two, one for pH titration in the range from 7 to 10.4 and the other for pH titration from 7 to 3.5. The pH values were measured with a Accumet XL25 pH-meter (Fischer Scientific, Fair Lawn, NJ) equipped with an NMR electrode, calibrated with pH 4.0, 7.0, and 10.0 standard solutions (Fischer Scientific, Fair Lawn, NJ). The pHs of the samples were adjusted with 0.1 or 0.2 M HCl or NaOH. $^1H-^{15}N$ HMQC spectra optimized for observing long-range $^1H-^{15}N$ correlation[39] were acquired with 2048 × 288 complex data points, spectral widths of 100 ppm in ^{15}N and 16 ppm in ^1H and 128 scans. The spectrum center was set as 200.0 ppm for ^{15}N and 4.76 ppm for ^1H. The observed ^{15}N chemical shifts of the histidine imidazole ring, provided that the full titration curves were available, were plotted against pH and fit with nonlinear least-squares regression analysis according to the Henderson-Hasselblach equations using the software Origin (Microcal Origain, Northampton, MA). The Henderson−Hasselbalch equation using a Hill parameter (n_H) to account for nonideality is given by:

$$\delta_{obs} = \frac{(\delta_{HA} + \delta_{A^-}10^{n_H(pH-pK_a)})}{(1 + 10^{n_H(pH-pK_a)})} \quad (1)$$

where δ_{obs}, δ_{HA}, and δ_{A^-} are the chemical shifts for the observed, protonated, and deprotonated species, respectively. The pK_a values for ΔΔIhh-CM and NI were determined in 100% D_2O samples while pK_a values for ΔΔIhh-SM were determined in 10% D_2O samples. The pK_a values were corrected for D_2O according to Krezel et al.[40]

pK_a Estimation with Limited Number of Titration Points. The average chemical shift of histidine ring nitrogens is a good indication of its ionization state.[41−43] When obtaining the full titration curve was not practical, the pK_a of histidines in NI was estimated by using the following equation when the neutral histidine is in fast tautomer exchange:

$$pK_a = pH + \log\left(\frac{(\tilde{A}_H - \tilde{A}_{obs})}{(\tilde{A}_{obs} - \tilde{A}_{H^+})}\right) \quad (2)$$

where $\tilde{A}_{obs} = 1/2(\delta_{\delta(obs)} + \delta_{\epsilon(obs)})$, $\delta_{\delta(obs)}$ and $\delta_{\epsilon(obs)}$ being the ^{15}N chemical shifts of histidine $N^{\delta1}$ and $N^{\epsilon2}$ atoms measured by HMQC NMR spectrum at individual pH values; $\tilde{A}_H = 1/2(\delta_\delta + \delta_{\epsilon H}) = 1/2(\delta_\epsilon + \delta_{\delta H})$, δ_δ, δ_ϵ, $\delta_{\delta H}$, and $\delta_{\epsilon H}$ are the ^{15}N chemical shifts of histidine (with either δ or ϵ position protonated) when the imidazole ring is neutral; $\tilde{A}_{H^+} = 1/2(\delta_{\delta H}{}^+ + \delta_{\epsilon H}{}^+)$, $\delta_{\delta H}{}^+$, and $\delta_{\epsilon H}{}^+$ is the ^{15}N chemical shifts of histidine with both δ and ϵ position protonated in a positively charged imidazole ring. The derivation of the equation and details of pK_a estimation are provided in the Supporting Information.

QM/MM Modeling of Reaction Mechanism. The *Mtu* recA intein splicing mutant (SM) crystal structure (PDB code 2IMZ)[30] was computationally appended to include both N- and C-exteins. The N-extein sequence consisted of Acetyl-VVKNK and the C-extein sequence consisted of CSPPF-N-methyl, both based on the native extein sequence. AMBER force field parameters[44] were implemented with GROMACS code.[45] This system was solvated with water molecules and equilibrated with classical molecular dynamics simulations with exteins present. MD simulations were carried out for 4 ns (0.5 ns equilibration, 3.5 ns production run) with temperature $T = 298$ K, pressure = 1 bar.

To test the N−S acyl shift and H73 acid/base behavior, multiscale methods were implemented. Specifically, hybrid quantum mechanical and molecular mechanical (QM/MM) calculations[46,47] were performed with Gaussian code,[48] using the B3LYP functional[49] with 6-31G(d,p) basis sets[50] for the quantum region and the AMBER force field parameters for the classical region. This multiscale method allows for bond breakage/formation within the QM active site while the remaining protein and solvent creates the structural backbone based on the folded protein and the electrostatic environment based on classical point charges. The protein system

(37) Du, Z.; Liu, Y.; McCallum, S. A.; Dansereau, J. T.; Derbyshire, V.; Belfort, M.; Belfort, G.; Van Roey, P.; Wang, C. *Biomol. NMR Assignments* **2008**, *2*, 111–113.
(38) Yamazaki, T.; Forman-Kay, J. D.; Kay, L. E. *J. Am. Chem. Soc.* **1993**, *115*, 11054–11055.
(39) Bax, A.; Summers, M. F. *J. Am. Chem. Soc.* **1986**, *108*, 2093–2094.
(40) Krezel, A.; Bal, W. *J. Inorg. Biochem.* **2004**, *98*, 161–6.
(41) Bachovchin, W. W. *Magn. Reson. Chem.* **2001**, *39*, S199−S213.
(42) Bachovchin, W. W.; Roberts, J. D. *J. Am. Chem. Soc.* **1978**, *100*, 8041–8047.
(43) Farr-Jones, S.; Wong, W. Y. L.; Gutheil, W. G.; Bachovchin, W. W. *J. Am. Chem. Soc.* **1993**, *115*, 6813–6819.
(44) Cornell, W. D.; Cieplak, P.; Bayly, C. I.; Gould, I. R.; Merz, K. M., Jr.; Ferguson, D. M.; Spellmeyer, D. C.; Fox, T.; Caldwell, J. W.; Kollman, P. A. *J. Am. Chem. Soc.* **1995**, *117*, 5179–97.
(45) Van Der Spoel, D.; Lindahl, E.; Hess, B.; Groenhof, G.; Mark, A. E.; Berendsen, H. J. C. *J. Comput. Chem.* **2005**, *26*, 1701–1718.
(46) Vreven, T.; Morokuma, K.; Farkas, O.; Schlegel, H. B.; Frisch, M. J. *J. Comput. Chem.* **2003**, *24*, 760–769.
(47) Maseras, F.; Morokuma, K. *J. Comput. Chem.* **1995**, *16*, 1170–9.
(48) Frisch, M. J.; et al. *Gaussian 03, Revision C.02*; Gaussian, Inc.: Wallingford, CT, 2004.
(49) Becke, A. D. *J. Chem. Phys.* **1993**, *98*, 5648–52.
(50) Hehre, W. J.; Ditchfield, R.; Pople, J. A. *J. Chem. Phys.* **1972**, *56*, 2257–61.

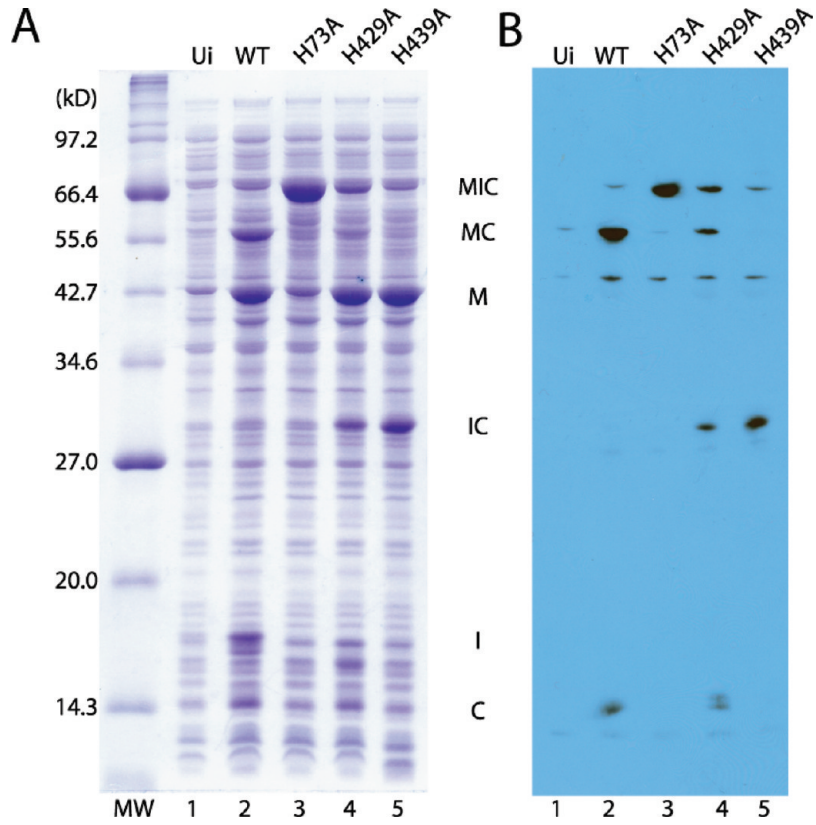

Figure 4. Inhibition of protein splicing by mutations of conserved histidines (B-block histidine H73A, F-block histidine H429A and penultimate histidine H439A). (A) SDS-PAGE of *in vivo* splicing assay. (B) Western blot with an anti-C-extein antibody. MW, molecular weight marker; Ui, uninduced; MIC, MBP-intein-C-extein fusion (71.5 kD); MC, MBP-C-extein fusion (56.1 kD); M, MBP (43.0 kD); IC, intein-C-extein fusion (28.6 kD) I, intein (15.4 kD).

consisted of 2351 atoms and there were 1321 water molecules present, for a system total of 6314 atoms. The quantum mechanical region is composed of the following residues: The N-terminal active site is based on the N-extein residue K(−1) and the intein residues C1-L2. Also included are T70, H73, and D422. From the G block and C-extein (C-terminal active site): V438−H439−N440-C(+1)-S(+2). At least three vicinal water molecules are included in the active site QM region. For nonmechanistic residues, those that affect the QM system via polarization, the entire amino acid may not be chosen for inclusion in the QM system: an example of such a residue is T70, where only the side chain is included. Similar density functional theory (DFT) methods have been tested in our own work[51] and that of others[52] and are shown to provide accurate energetic and structural results.

Results and Discussion

B-Block Histidine H73 Is Essential for Protein Splicing in ΔΔIhh-SM. To show the functional significance of the B-block histidine H73 in minimized and engineered intein ΔΔIhh-SM, we tested the effect of an H73A mutation on splicing using an *in vivo* splicing assay. *E.coli* cells were induced to express an intein precursor MIC, composed of maltose binding protein (MBP) as N-extein, ΔΔIhh-SM intein and a short C-extein. The precursor underwent splicing, N- and C-terminal cleavage (Figure S1, Supporting Information) within the cells and the extent of splicing was assessed with SDS-PAGE and Western blot using an antibody against the C-extein (Figure 4). With the WT ΔΔIhh-SM as the intein (H73), ligated exteins (MC) and large amount of I (intein) was generated by protein splicing and there was little presplicing precursor MIC remaining. Products of N-terminal cleavage (M and IC) and C-terminal cleavage (C) were also observed. Thus, the engineered and minimized intein ΔΔIhh-SM can efficiently catalyze protein splicing and is a good model system for exploring the structural mechanism of protein splicing at atomic resolution.

Next we examined the effect of alanine mutation of the conserved histidines. In contrast to the wild type, in the H73A mutant, the splice products were completely absent and there was accumulation of large amount of MIC precursor (Figure 4), indicating an essential role for B-block histidine in protein splicing. H73A also caused dramatic reduction in N-terminal cleavage product (M) (Figure 4, lanes 2 and 3), consistent with the notion that the B-block histidine plays a critical role in N-X acyl shift.[18,25,26]

The H429A mutation greatly reduced the amount of splice product (Figure 4B, lanes 2 and 4), indicating an important but nonessential role for the F-block histidine in splicing. Large amount of N-terminal cleavage product M was still produced with H429A, suggesting F-block histidine does not contribute critically to N-terminal cleavage. The H429A mutation caused a reduction in the amount of C-terminal cleavage product C compared with WT, suggesting F-block histidine likely plays a role in C-terminal cleavage (Figure 4B).

No splice product MC was generated with the H439A mutation, demonstrating the essential role of the penultimate histidine in protein splicing (Figure 4A and B, lane 5). However, very little precursor protein MIC was observed with the H439A

(51) Shemella, P. Ph.D. Disssertation, Renssealer Polytechnic Institute, 2008.
(52) Elstner, M.; Frauenheim, T.; Suhai, S. *THEOCHEM* **2003**, *632*, 29–41.

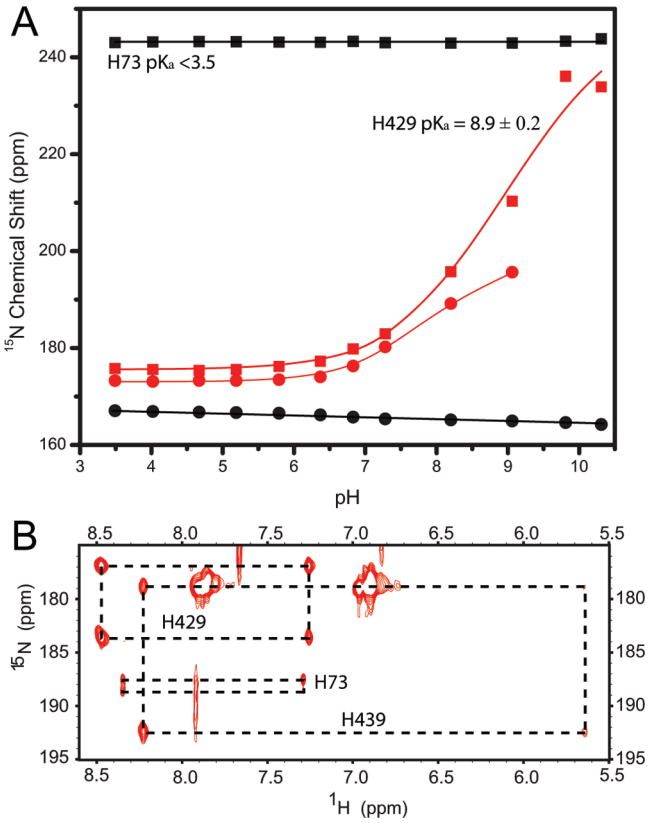

Figure 5. pK_a determination of conserved histidines in spliced intein (A) and in a precursor with an N-extein, NI (B). (A) Plot of the ring ^{15}N chemical shift versus pH for conserved histidines, B-block H73, F-block H429 in ΔΔIhh-SM. The squares and circles represent the chemical shifts of $N^{\delta 1}$ and $N^{\epsilon 2}$, respectively. The pK_a of H439 in the spliced intein is shown in Table 1 along with pK_a's of other histidines in different constructs. (B) HMQC spectrum for intein precursor NI at pH 7.1. The average chemical shift of of H73 $N^{\delta 1}$ and $N^{\epsilon 2}$ is 188 ppm, midway between the values expected for neutral (204.3 ppm) and charged histidines (176 ppm), suggesting that H73 has a pK_a near 7 in the precursor protein.

mutation. While large amount of N-terminal cleavage product M was produced, no C-terminal cleavage product C was generated. Significant amount of IC was accumulated without undergoing further C-terminal cleavage. These observations indicated that H439A blocked C-terminal cleavage.

Thus, not only is the B-block histidine the most conserved among three histidines, but also its mutation (H73A) causes the most dramatic inhibition of protein splicing and cleavage reactions.

B-Block Histidine Has a Strikingly Low pK_a in Spliced Intein Product. To explore the mechanistic role of the B-block histidine in protein splicing, we examined the ionization state of H73 in the spliced intein product ΔΔIhh-SM. Aromatic nitrogen resonances of H73 did not titrate in the experimental pH range of 3.5 to 10 (Figure 5A), indicating that the pK_a of H73 is significantly less than 3.5 and that H73 is a relatively strong acid in the intein product. It is known that acidic conditions favor the N−S (N−O) acyl shift in peptides.[53] H73 may therefore facilitate the N-X acyl shift in the first step of protein splicing as a general acid to protonate the leaving nitrogen group. This catalytic role for the B-block histidine has been well-recognized[7] and is supported by the short distance

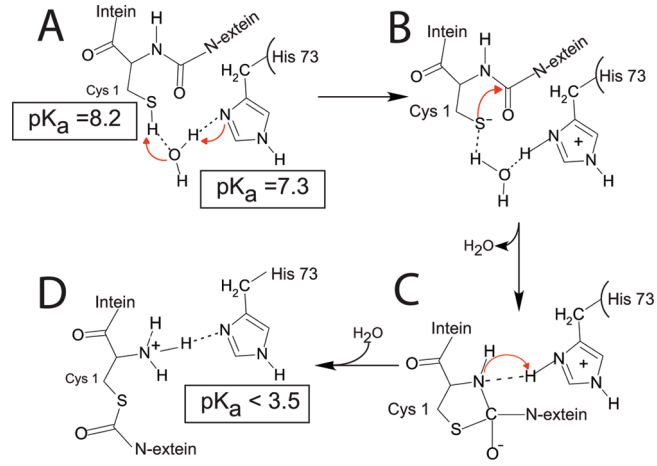

Figure 6. pK_a shift mechanism for B-block histidine H73 in the N−S acyl shift, the first step of protein splicing for the *Mtu* RecA intein. H73 acts as a general base to deprotonate the C1 thiol, initiating the N−S acyl shift. Then H73 serves as a general acid to protonate the leaving nitrogen, breaking the scissile bond and completing the N−S acyl shift. The arrows indicate the routes of electron transfer and broken lines indicate hydrogen bonds. Intermediates A−D are described in the text.

Table 1. pK_a Values of Histidines in Various Intein Constructs

histidine pK_a	ΔΔIhh-SM	ΔΔIhh-CM	NI
B-block H73	<3.5	<4.1	7.3 ± 0.6
F-block H429	8.9 ± 0.2	8.4 ± 0.5	7.9 ± 0.6
G-block H439	6.2 ± 0.2	6.2 ± 0.2	7.3 ± 0.6
His 17	4.5 ± 0.3	4.7 ± 0.2	4.4 ± 0.6
His 30	7.1 ± 0.2	6.6 ± 0.4	6.6 ± 0.6
His 41	6.7 ± 0.2	6.3 ± 0.2	6.6 ± 0.6

between the H73 ring and C1 amide nitrogen in the crystal structures of the *Mtu* RecA (Figure 3) and other inteins.[17,19,27−29] The depressed pK_a for the B-block histidine is consistent with the essential role that this residue plays in the breakdown of the scissile bond at the N-terminal splicing junction, by acting as a general acid (Figure 6).

Two factors could contribute to the low pK_a of H73 in ΔΔIhh-SM: *First*, since H73 is 97% buried from solvent (calculated by the program molmol[54]), the energetic cost of a charged histidine ring could be high due to the low dielectric constant in the interior of a protein. However, we have evidence that the low pK_a of H73 is not due to the inaccessibility of solvent. Although H73 appears buried in the static crystal structure, the hydrogen−deuterium exchange was complete within minutes for the H73 backbone amide and the ϵ2 protons. Therefore the H73 ring has adequate access to solvent pH conditions through protein dynamics. *Second*, since H73 is proximal to the positively charged N-terminus (2.9 Å between H73 $N^{\delta 1}$ and C1 amide N in the crystal structure; Figure 3), a positively charged histidine ring will cause an unfavorable electrostatic interaction. This second factor lead us to hypothesize that in the absence of a nearby charged N-terminus in precursors with an N-extein, H73 is more likely to become protonated and should have a higher pK_a. Thus, in a precursor, the higher pK_a may enable the B-block histidine to act as a base, to deprotonate the C1 thiol group to initiate the first step of splicing (Figure 6). We call this the pK_a shift hypothesis in which the B-block histidine

(53) Iwai, K.; Ando, T.; Hirs, C. H. W. *Methods in Enzymology*; Academic Press: New York, 1967; Vol. 11, p 263.

(54) Koradi, R.; Billeter, M.; Wuthrich, K. *J. Mol. Graph.* **1996**, *14*, 51–5, 29−32.

acts as a general base first to initiate N-X acyl shift and then serves as a general acid to complete the N-X acyl shift (Figure 6).

B-Block Histidine Has a pK_a of 7.3 in Presplicing Precursor. To test our pK_a shift hypothesis, we estimated the pK_a of H73 in an intein precursor NI, composed of a short N-extein and ΔΔIhh-SM sequence with a C1A mutation (Figure 5B) to prevent N-terminal cleavage. It is possible that the C1A mutation might modify H73 pK_a in NI. However, as noted above, C1 influenced the pK_a of H73 mainly through its amide group. Therefore mutating C1 side chain was unlikely to dramatically change H73 pK_a in the precursor. In NI, full pH titration was not possible because NI precipitated at pH > 8 and aggregated at pH < 4 which resulted in severe line-broadening. When the histidine tautomer exchange is fast, the average chemical shift of $N^{\delta 1}$ and $N^{\epsilon 2}$ reflects the effect of the pK_a equilibrium.[41−43] We therefore estimated the pK_a of H73 in the precursor using the following equation:

$$pK_a = pH + \log\left(\frac{(\tilde{A}_H - \tilde{A})}{(\tilde{A} - \tilde{A}_{H+})}\right)$$

where $\tilde{A}$ is the average $N^{\delta 1}$ and $N^{\epsilon 2}$ chemical shift measured from HQMC at a pH value close to pK_a; $\tilde{A}_H$ and $\tilde{A}_{H+}$ are the average $N^{\delta 1}$ and $N^{\epsilon 2}$ chemical shift of the neutral imidazole ring and of the charged imidazolium ring, respectively. $\tilde{A}_H$ was calculated from the data in Figure 5A to be 204.3 ppm, while $\tilde{A}_{H+}$ was taken from the literature as 176 ppm.[55] This method for pK_a estimation has been validated using the pK_a titration data in ΔΔIhh-SM where full titration curves are available (see Supporting Information). Using this method, we obtained a pK_a value of 7.3 ± 0.6 for H73 in NI precursor. This result demonstrates a dramatic pK_a shift, of at least 4 units, for the B-block histidine during protein splicing.

pK_a Shift Mechanism for B-Block Histidine in Protein Splicing. The pK_a value of the C1 thiol group has been determined by Paulus et al. to be 8.2,[56] within 1 pH unit of the pK_a values of H73 (7.3) in an intein precursor. Thus a proton transfer from the thiol group to the imidazole ring can occur at a reasonable rate.[57] Given the proximity of their side chains to the C1 thiol group (within 6 Å) in the crystal structure of ΔΔIhh-SM, D422 and H73 are the most likely candidates to act as a general base to activate the C1 thiol group. However, the D422 side chain is most likely to have a pK_a near 4 in the intein precursor and would be much less efficient as a general base compared with H73 because of the large pK_a difference between D422 and C1. Indeed, various mutations of D422 can still mediate N-terminal cleavage,[58] suggesting D422 is not essential for the N−S acyl shift. In addition, D422 is not as highly conserved as the B-block histidine (Figure 2). Thus, in the intein precursor, H73 is most likely the general base to deprotonate the C1 thiol to initiate the first step of splicing, the N−S acyl shift (Figure 6). The distance between H73 and C1 is 5.7 Å in the crystal structure, too great for direct proton transfer. However, the proton transfer can occur through an intermediate water molecule. Indeed, a water molecule forming hydrogen bonds with both the C1 thiol group and the H73 ring nitrogen has been observed in our simulation with a precursor based on the crystal structure of ΔI-SM (Figure 6A). However, such a water molecule has not been observed in any crystal structures of intein precursors. This could be due to mutations such as C1S or H73N in the precursor sequence used for crystallization or the intrinsic low sensitivity for observing water molecules by crystallography. Alternatively, a conformational change may occur in the intein precursor to bring the H73 $\delta 1$ nitrogen close to the C1 thiol group for a direct activation.

After the proton transfer between the C1 thiol group and H73, the positively charged H73 ring stabilizes the negatively charged C1 thiolate and oxythiazolidine intermediate (Figure 6B). H73 will then protonate the nascent N-terminal amide group, causing the breakdown of the scissile peptide bond (Figure 6C). An additional proton from water generates a positive charge at the intein N-terminus and lowers the pK_a of H73 to below 3.5, stabilizing the endproduct of N−S acyl shift (Figure 6D). In this pK_a shift mechanism, H73 acts both as the catalytic base and acid for the N−S acyl shift, supported by the large pK_a shift observed here. Similar pK_a shifts associated with a dual role of a catalytic residue in acid/base catalysis have been observed in β-glucosidase.[59]

QM/MM Modeling Supports the pK_a Shift Mechanism of the B-Block Histidine. Combined QM/MM modeling was applied to further evaluate the pK_a shift mechanism in protein splicing. QM/MM is the method of choice for modeling chemical reactions catalyzed by enzymes.[60] QM is required to describe the electron and proton transfers at the active site. However, the properties of the protein scaffold and solvent have a strong effect on enzyme reaction mechanism. For example, the pK_a of a well-solvated histidine on the protein surface will be different from the pK_a of a histidine buried in the hydrophobic core. Therefore, atoms outside of the active site are modeled with classical MM. Atomic level modeling of the autocatalytic splicing reaction of inteins involves understanding the role of protons as a function of tertiary structure. The protein backbone is semirigid, and with multiscale quantum and classical modeling (QM/MM), the structural and electrostatic environment can be a perturbation on the quantum mechanical active site. Currently full-protein QM simulations are computationally unfeasible, and with QM/MM multiscale methods we can model the position of protons and their catalytic role, which is intimately related to the pK_a measurements presented in this work.

Multiscale QM/MM was used to create an energy profile for C1 ionization, N−S acyl shift, and for protonation of the nascent N-terminal amide (Figure 7). The QM/MM modeling of the intein's N-terminal rearrangement reaction corroborates the dual catalytic role of H73. H73 was first used as a catalytic base, removing a proton from the C1 thiol group via water, and then as an acid, by protonating the leaving nitrogen. The ionization energy for C1 by H73 via water was ~19.2 kcal/mol, comparable to the energy barrier of C-terminal cleavage in inteins.[51,61] Once the proton has migrated from C1 to H73, the C1 thiolate group is activated to attack the peptide carbonyl carbon, forming a transitory ring structure, oxythiazolidine, with an energy 12 kcal/mol higher than the ionized C1 state. The oxythiazolidine intermediate is highly activated for N-protonation and C−N

(55) William, W. B. *Magn. Reson. Chem.* **2001**, *39*, S199−S213.
(56) Shingledecker, K.; Jiang, S.-q.; Paulus, H. *Arch. Biochem. Biophys.* **2000**, *375*, 138–144.
(57) Silverman, D. N.; Tu, C.; Chen, X.; Tanhauser, S. M.; Kresge, A. J.; Laipis, P. J. *Biochemistry* **1993**, *32*, 10757–62.
(58) Pereira, B. Ph. D. Disssertation, Renssealer Polytechnic Institute, 2008.
(59) McIntosh, L. P.; Hand, G.; Johnson, P. E.; Joshi, M. D.; Korner, M.; Plesniak, L. A.; Ziser, L.; Wakarchuk, W. W.; Withers, S. G. *Biochemistry* **1996**, *35*, 9958–66.
(60) Senn, H. M.; Thiel, W. *Angew. Chem., Int. Ed.* **2009**, *48*, 1198–229.
(61) Shemella, P.; Pereira, B.; Zhang, Y.; Van Roey, P.; Belfort, G.; Garde, S.; Nayak, S. K. *Biophys. J.* **2007**, *92*, 847–53.

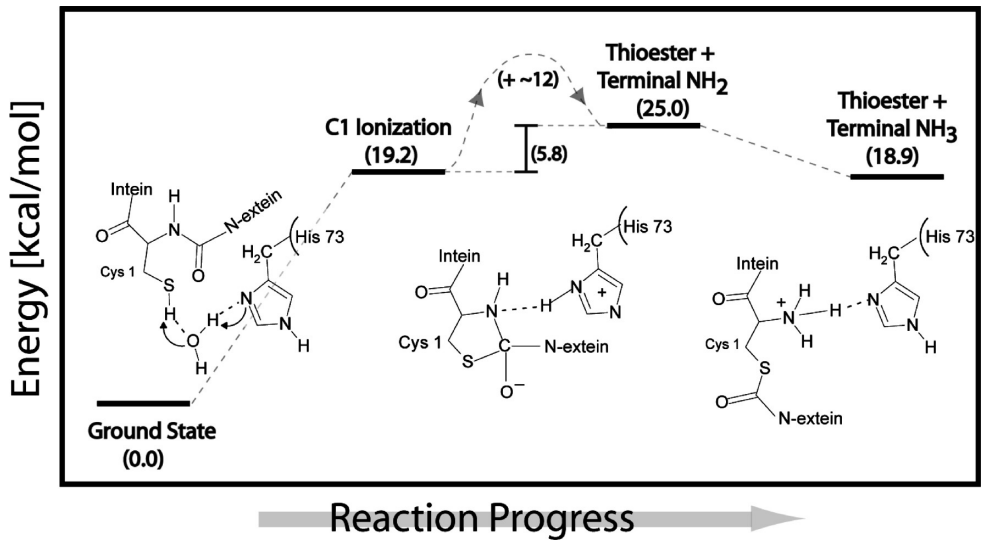

Figure 7. Reaction profile for N−S acyl shift derived from QM/MM simulations.

bond breakage. The newly formed C−S bond caused a decrease in resonance at the peptide bond and an increased bond length between C and amide N, resulting in a more nucleophilic scissile bond amide N. The recently protonated H73 was in position to directly and spontaneously donate a proton to the scissile peptide bond nitrogen, without the need for a transitory water molecule. The energy of the thioester with an -NH_2 terminal group was 25.0 kcal/mol, and it was determined to be a structurally stable intermediate. A second protonation then occurred at the N-terminus, via solvent, resulting in a positively charged -NH_3^+ group that was electrostatically coupled to the neutral H73 (with energy 18.9 kcal/mol). Indeed, the average pK_a for a solvent exposed N-terminus ranges from 8.8 to 10.8.[62] The positive N-terminus considered in QM/MM simulations caused H73 to be energetically unable to accept a second proton, consistent with the extreme reduction of the H73 pK_a in the intein product observed experimentally. The QM/MM simulations of the C1 ionization, the N−S acyl shift, and the N-terminal thioester formation are fully consistent with the observed pK_a values of H73 before and after N−S acyl shift.

Mechanistic Implications of pK_a Values of Other Conserved Histidines. The conserved F-block histidine, H429 in *Mtu* recA, showed a higher than usual pK_a in both the spliced product and precursor NI (Table 1), which suggests that H429 may serve as a proton acceptor in the splicing reaction. Indeed, the F-block histidine was shown to play an important role in protein splicing, especially in C-terminal cleavage (Figure 4).[12] The crystal structure of ΔΔIhh-SM[30] showed that the imidazole ring of H429 can form a hydrogen bond with a water molecule, Wat15, which in turn has a hydrogen bond with the NH_2 side chain of N440. The F-block histidine may deprotonate the N^{δ} atom of N440 through this hydrogen bond network to activate the N^{δ} of N440, which then carries out the nucleophilic attack on the carbonyl of the scissile bond, resulting in asparagine cyclization and the breakdown of the scissile bond, as suggested by Sun et al. in Ssp DnaE intein.[63] The pK_a of H439 was determined to be ~6, indicating that the H439 ring is neutral at pH 7. H439 may act as a hydrogen bond donor to the scissile bond carbonyl, as shown by the crystal structure of ΔΔIhh-SM (Figure 8).[30]

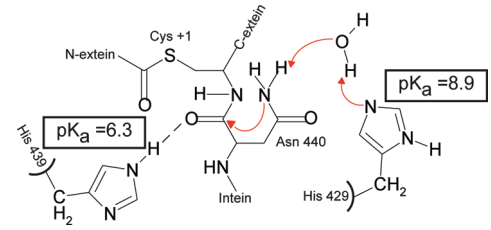

Figure 8. Mechanistic role of the F-block histidine and penultimate histidine.

Histidine pK_a values are similar in ΔΔIhh-SM and the cleavage mutant ΔΔIhh-CM (Table 1); therefore the difference in splicing and cleavage activities between these two mutants is not due to the differences in the ionization state of histidines in the spliced product. As far as other histidines are concerned, H17 has a low pK_a of ~4.5, likely because it is 89% buried from the solvent (calculated by molmol[54] using the crystal structure of ΔΔIhh-SM). H30 and H41 have normal pK_a values, which are the same in the precursor and the product. Thus, the nonconserved H17, H30 and H41 do not have distinctive roles in splicing, while the B-, F-, and G-block histidines have important functions in intein catalysis with their characteristic pK_a values.

Conclusions

In summary, we demonstrate a dramatic pK_a shift for the B-block histidine, the most conserved intein residue, during protein splicing. The combined NMR and QM/MM data suggest that the B-block histidine has a dual role in acid−base catalysis of protein splicing. The B-block histidine likely acts as a general base to initiate N−X acyl shift then as a general acid to cause the breakdown of the scissile bond. The proposed pK_a shift mechanism accounts for the absolute sequence conservation of the B-block histidine, and is in accord with biochemical data supporting the critical role of the B-block histidine in the N−X acyl shift and splicing. The proposed mechanism will likely have important implications for the development of intein-based biotechnologies for protein purification, protein engineering and

(62) Voet, D.; Voet, J. G. *Biochemistry*, 3rd ed.; John Wiley and Sons: New York, 2004.
(63) Sun, P.; Ye, S.; Ferrandon, S.; Evans, T. C.; Xu, M.-Q.; Rao, Z. *J. Mol. Biol.* **2005**, *353*, 1093–1105.

biosensing. Finally, it is important to note that proposed mechanism is likely one plausible mechanism for protein splicing rather than the only way in which splicing reaction operates.

Acknowledgment. We thank financial support from National Institute of Health (R01GM81408 to C.W.; R01GM44844 to M.B.), the National Science Foundation (CTS03-04055-NIRT to G.B.), and the Computational Center for Nanotechnology Innovations (CCNI) at RPI. P.T.S. acknowledges funding from the Interconnect Focus Center. Y.L. acknowledges the Cultivation Fund of the Key Scientific and Technical Innovation Project, Ministry of Education of China (No. 707036). B.P. acknowledges funding from NIH Biomolecular Science and Engineering training grant (GM067545). We thank Shekhar Garde and Patrick Van Roey for helpful discussions.

Supporting Information Available: One table describing all the mutations used in this paper (Table S1), validation of the pK_a determination using limited titration points (Table S2), complete ref 48, and four figures (protein splicing and its side reactions, and assignment of histidine resonances in ΔΔIhh-SM, ΔΔIhh-CM, and NI). This material is available free of charge via the Internet at http://pubs.acs.org.

JA904318W